

REVIEW

Open Access



# Metagenomic applications in exploration and development of novel enzymes from nature: a review

Fitra Adi Prayogo<sup>1\*</sup>, Anto Budiharjo<sup>2,3\*</sup>, Hermin Pancasakti Kusumaningrum<sup>1</sup>, Wijanarka Wijanarka<sup>2</sup>, Agung Suprihadi<sup>2</sup> and Nurhayati Nurhayati<sup>2</sup>

## Abstract

**Background:** Microbial community has an essential role in various fields, especially the industrial sector. Microbes produce metabolites in the form of enzymes, which are one of the essential compounds for industrial processes. Unfortunately, there are still numerous microbes that cannot be identified and cultivated because of the limitations of the culture-based method. The metagenomic approach is a solution for researchers to overcome these problems.

**The main body of the abstract:** Metagenomics is a strategy used to analyze the genomes of microbial communities in the environment directly. Metagenomics application used to explore novel enzymes is essential because it allows researchers to obtain data on microbial diversity, reaching of 99% and various types of genes encoding an enzyme that has not yet been identified. Basic methods in metagenomics have been developed and are commonly used in various studies. A basic understanding of metagenomics for researchers is needed, especially young researchers to support the success of the research.

**Short conclusion:** Therefore, this review was done in order to provide a deep understanding of metagenomics. It also discussed the application and basic methods of metagenomics in the exploration of novel enzymes, especially in the latest research. Several types of enzymes, such as cellulases, proteases, and lipases, which have been explored using metagenomics, were reviewed in this article.

**Keywords:** Metagenomics, Novel enzymes, Microbial community, Environmental DNA

## Background

The microorganism community from nature is the largest community that plays an essential role in the biogeochemical cycle on earth. Many microorganisms are also known to have a role in the development of the industry that exists today by the production of metabolites [1]. Enzymes are one of the microbial metabolites often used in the industrial processes.

Enzymes are biocatalyst compounds that can accelerate biochemical reactions used in various industries, such as

textiles, paper, detergents, food, and beverages [2]. Various benefits of enzymes have attracted the attention of researchers to develop and explore enzymes from nature for further application in the industrial field. Unfortunately, there are still many types of microorganisms that are not identified yet and cannot be cultured in the growth media. The use of culture-based method only results in diversity data of less than 1% of the total microorganisms in the environment [3].

Metagenomics is a breakthrough for the weakness of culture-based method, which has sharply increased its application in recent years. In the metagenomics, DNA is directly extracted from the environment samples without culturing process in the laboratory. The use of DNA to analyze the diversity of microorganisms reveals a representative and comprehensive result [4, 5]. Metagenomics has been used in

\* Correspondence: [fitra.adiii@student.undip.ac.id](mailto:fitra.adiii@student.undip.ac.id);  
[anto.budiharjo@fulbrightmail.org](mailto:anto.budiharjo@fulbrightmail.org)

<sup>1</sup>Department of Biology, Faculty of Science and Mathematics, Diponegoro University, Semarang City 50275, Indonesia

<sup>2</sup>Biotechnology Study Program, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Sudharto SH, Semarang 50275, Indonesia  
Full list of author information is available at the end of the article

various fields of study, such as in the microbial communities of the human intestine [6], sugarcane bagasse waste [7], and hypersaline environment [8]. In addition to exploring the benefits of gene resources from nature, the existence of metagenomics studies can also increase knowledge about the relationships between microorganism communities in the biogeochemical cycle in nature.

The understanding of metagenomics needs to be reviewed further in order to deepen the insights of metagenomic studies. A thorough understanding of metagenomics and their application in research is expected to have an impact on increasing discoveries about the information of the microbial community and enzymes from nature. Therefore, this review is designed to discuss the application of metagenomics in the exploration of novel enzymes from nature. The focus of this review is to provide a deep understanding of metagenomics, basic method, and its utilization to enzyme exploration, especially in the latest research.

## Main text

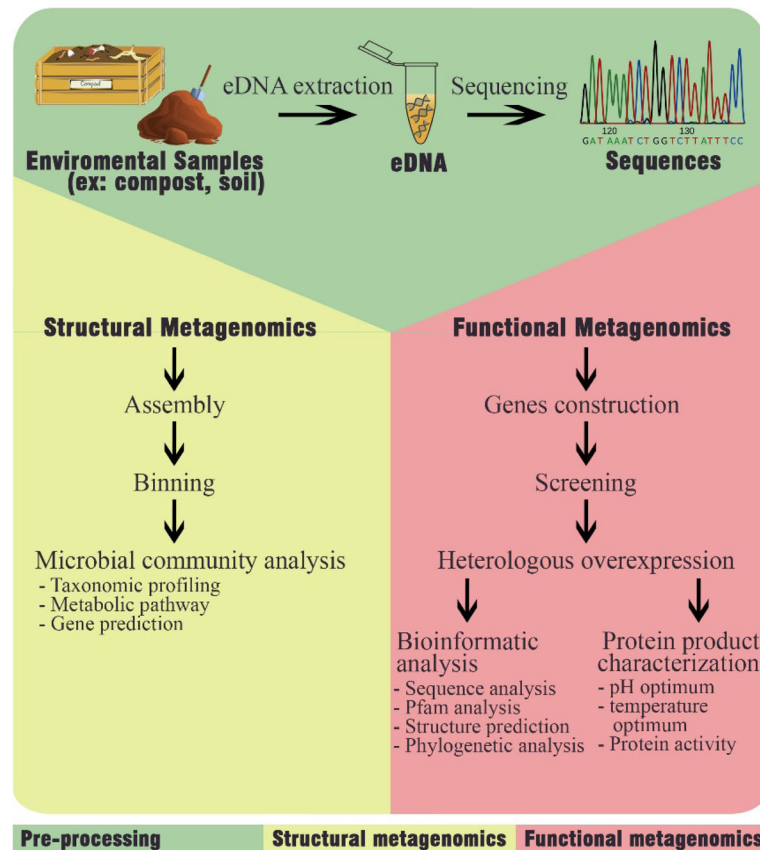
### Metagenomics

Direct DNA extraction from the environment was started in 1985 by Pace and his team. However, the new term of metagenome emerged in 1998 by a researcher

named Handelsman. Metagenomics is the study of genomes from microorganism communities in the environment [9, 10]. Other terms of metagenomics are community genomics, environmental genomics, and population genomics [4]. Metagenomics is a strategy used to analyze genomes acquired from the community of environmental microorganisms without culturing them [11]. This technique can read the diversity of microorganisms up to 99% of the total microorganisms in environmental samples [12]. Metagenomics becomes a new concept in microbiology studies, thus opening the horizons of researchers' minds to discover new biochemical compounds that are available in nature and can be utilized in the biotechnology industry.

### The direction of metagenomics study

Figure 1 shows the direction in a metagenomics study. Metagenomics is divided into two primary studies, namely, structural metagenomics and functional metagenomics [13]. Structural metagenomics is a study focused on the structure of microbial communities. The study of community structure focuses on understanding the relationships between individual components in building a community in an environment. Relationships between



**Fig. 1** Framework for metagenomics with two primary studies, structural and functional metagenomics

components in the community are essential information for studying ecology and biological functions [12]. Basic structural metagenomics methods consist of assembly, binning, and microbial community analysis such as taxonomic profiling, gene prediction, and metabolic pathways [3, 14].

Functional metagenomics is a study focused on the use of genes encoding a particular protein [12]. The study of functional metagenomics is a new challenge in exploring natural compounds that can be utilized in the biotechnology industry. Several basic methods in functional metagenomics are carried out to access the novel enzymes, like gene construction, screening, gene expression, and can be followed by bioinformatic analysis such as sequence, Pfam, structure prediction, and phylogenetic analysis and also protein product characterization such as optimum pH rate, optimum temperature rate, and protein activity analysis [10].

The two approaches, structural and functional metagenomics, are a strategy for the exploration of microorganism communities in ecology and biotechnology studies. This combination cannot be separated in metagenomic studies. Both are the basis of microbial ecological problems, namely, "What types of microorganisms exist in the environment?" Furthermore, "What is the function of these microorganisms in the environment?" [12].

#### Microbial community as metagenomic research object

The microbial community is the largest community that plays an essential role in the biogeochemical cycle on the planet [1]. Microbial communities have the most diverse species on earth by forming 60% of the earth's biomass [15]. The total number of microbes on earth is even predicted to reach  $10^{30}$  [16]. The importance of the role of microbes in the "balance of life" on this planet makes us need to deepen understanding of the microbial community, so that ecosystem damage does not occur. Better ecosystem management and progress in bioprospection will be achieved with a fundamental understanding of interactions between microbial communities [1]. Microbial community with various biochemical reactions in it is a mystery that is still a question mark for researchers. The existence of metagenomics allows researchers to unravel the mysteries that are in it. The microbial community will be something interesting to continue to study.

Furthermore, the microbial community also has benefits in industrial processes. Various types of enzymes found in several publications come from the microbial community, such as cellulases [17], proteases [18], and esterases [19]. Enzymes originating from the microbial community have advantages for industry compared to animals and plants, such as being more stable, have high yields, and are easily engineered [20].

#### Metagenomic sequencing technology

In the past, microbial analysis was done using pure culture. The use of pure culture by conventional isolation is a limiting factor in the analysis of environmental microbes. Microbial communities in the environment interact with each other to exchange nutrients, biochemical products, and chemical signals [21]. The presence of a microbial community complex system cannot be captured if it relies solely on a pure culture system.

The molecular method has made a new era in the analysis of microbial communities. Carl Woese started the concept of molecular analysis on microbes in the 1970s. He used rRNA as a molecular marker in classification analysis [22]. The use of sequencing at that time was conventional methods called Sanger [23]. The Sanger method is a sorting method that uses a single strand as a template. This method has the disadvantage of working for a long time and high running costs. Even the Sanger method will require approximately 15 years and cost the US \$ 100 million to do the sequencing of the human genome [24].

The second-generation sequencing method emerged after researchers used the Sanger method for more than three decades. This method is often referred to as next generation sequencing (NGS). Several technology platforms included in the second-generation sequencing method are Roche/454, Ion torrent, and Illumina [24]. According to Bragg & Tyson [25], second-generation sequencing has advantages over its predecessor, namely, (1) more efficient speed, (2) cheaper running costs, and (3) sequencing results that can be detected immediately without electrophoresis. Table 1 presents specific data on the different characteristics of the technology used in second-generation sequencing [23].

The second generation of sequencing technology that has been sufficiently developed still has problems regarding costs, results, and time that might be optimized again. Those problems lead to the development of third generation of sequencing technology. Third-generation sequencing has advantages over the second generation, namely, lower sequencing costs, no PCR processing, and a faster process [24]. The technology platforms included

**Table 1** Comparison of the characteristics in second-generation sequencing

Characteristics	Roche 454	Ion torrent	Illumina
Maximum read length (bp)	1200	400	300
One-way results (Gb)	1	2	1000
Amplification for library construction	Yes	Yes	Yes
Cost/Gb (\$)	9538.46	460.00	29.30
Error rate (%)	1	~ 1	~ 0.1
Running time (H)	20	7.3	144

in the third-generation sequencing method are PacBio RS (Pacific Bioscience) and Oxford Nanopore [23].

#### Basic methods in metagenomics analysis

Method selection is an essential strategy in the metagenomics analysis. In summary, the method is divided into two, namely, the molecular and bioinformatic methods [26].

#### Molecular method

Metagenomics is the study about the genome of the environmental community (metagenome) as the subject of research. This study is slightly different from genome studies focused on an individual (single genome).

**Metagenomic DNA extraction** The extraction of DNA metagenome is carried out directly from environmental samples. This process is the first step in accessing the DNA metagenome. Some researchers use different methods, depending on the type of research sample used [10]. Tanveer et al. [27] have carried out DNA extraction of the metagenome using commercial kits and standard protocols.

Metagenomic DNA extraction using commercial kits is the easiest method because it only uses chemicals that have been provided by the manufacturer. According to Lear et al. [28], some researchers use branded kits based on the type of sample to be extracted. The PowerSoil and DNeasy PowerMax (Qiagen) kits are the most popular kits for researchers in soil samples, while the DNeasy Blood and Tissue Kits (Qiagen) kits are the most commonly used kits for seawater and groundwater samples.

In contrast to commercial kits, the use of standard protocols takes longer time than commercial kits [29]. Therefore, researchers prefer kits because they are more efficient in terms of time. However, some studies that use standard protocols show better results when compared to kits. Tanveer et al. [27] tried to compare the extraction of metagenomic DNA from the soil using the HiPurA soil DNA isolation kit (Himedia) and standard protocol. The results revealed that the standard protocol produced the highest concentration of DNA. Hassan et al. [30, 31] have also proven that the use of standard protocols produces higher concentrations than the DNA isolation kit for water (Epicenter).

Metagenomic DNA extraction is a crucial process because it will have an impact on the success of the further stage. According to Felczykowska et al. [32], the extraction of metagenome must produce a perfect DNA size. The size of fragments typically used for metagenome analysis is 600 bp to 25 kbp. Poor results will make the extracted sample unusable for further metagenomic analysis. Therefore, it is necessary to pay attention to the following: (1) do not physically interfere with genetic material and (2) contamination with

protein, humic acid, and metals must be avoided. Other factors that might affect the results of DNA extraction are pH, soil mineral level, and soil type [33].

**Calculation of concentration and purity of metagenomic DNA extracts** Determination of DNA concentrations and purity values can be calculated through 3 methods, namely, UV absorbance, fluorescent staining, and diphenylamine reaction [34]. The UV absorbance method is the most popular method for researchers to calculate the concentration and purity of DNA. It is because the UV absorbance method is easy, practical, and inexpensive [35].

Calculating the concentration and purity of DNA requires a device known as a spectrophotometer [35]. The principle of the UV absorbance ray method is the utilization of specific wavelengths that can be captured by DNA molecules [34]. DNA has the highest UV absorption at a wavelength of 260 nm, while proteins at a wavelength of 280 nm. Therefore, the wavelength ratio used when calculating the purity of nucleic acids is A260/A280. DNA samples have a purity ratio of around 1.8–2.0 [35]. The ratio value 260/230 can be used to help evaluate the presence of salt compounds, proteins, guanidine HCL, EDTA, lipids, and phenols. The lower the value, the higher the number of contaminants [36].

Contaminants can worsen DNA purity results. The most common contaminants in metagenome samples are humic acid and protein [32]. Protein and phenol contaminants usually show absorption values of 260/280, which are lower than 1.6. Meanwhile, if the absorption ratio value of 260/280 is more than 2.0, it indicates the presence of RNA contamination to DNA [36].

**Gel Electrophoresis** Gel electrophoresis is a standard qualitative method used to separate, identify sizes, and purify nucleic acids. This method uses a gel media that has pores and can be passed through by nucleic acids [37]. Nucleic acids have phosphate groups that make these molecules negatively charged so that nucleic acid molecules will move towards the anode (positive electrode) when energized. The speed of this transfer is influenced by the factor of molecular weight, gel concentration, and the electrical voltage used [38].

Agarose gels are the most popular in gel electrophoresis. Agarose gels are polymers consisting of disaccharide units, which are arranged repeatedly and consist of galactose and 3,6-anhydrogalactose. This gel is made from seaweed extract and has large pores [37]. Pore size can be affected by gel concentration. Each gel concentration profile shows the optimal state of the length of the nucleic acid fragment used as a sample when running gel electrophoresis. Gutiérrez-lucas et al. [39] have used a 0.8% gel concentration for samples originating from



**Table 2** Recommended agarose gel concentrations based on fragment length from nucleic acid samples

Gel concentration (%)	The size of nucleic acids (kb)
0.3	5–60
0.6	1–20
0.7	0.8–10
0.8	1–7
0.9	0.5–7
1.2	0.4–6
1.5	0.2–3
2.0	0.1–2

the soil. The choice of 0.8% agarose gel concentration is a strategy for electrophoresis from metagenomic samples because environmental DNA fragments (eDNA) have an extended size. Table 2 presents recommendations for gel concentrations used and adjusted based on the length of the nucleotide acid fragments used for the sample [38]:

**Amplification of 16S rRNA gene** Ribosomes are essential compounds for protein synthesis. They are very conservative and often used as a standard for determining taxonomies. Prokaryotic microbes are generally composed of 65% rRNA (ribosome-ribonucleic Acid) and 35% protein. Each prokaryotic ribosome consists of 2 subunits, namely, large subunits (LSU) (the 50S), which contain two rRNA molecules (5S and 23S) and small subunits (SSU) (30S) that contain a single rRNA molecule (16S) [40].

16S rRNA is an area often used as a standard for taxonomy profiling analysis in prokaryotic organisms [41]. This gene has nine regions called hypervariable regions (V1–V9) with a total length of about 1500 bp. These nine regions can distinguish the diversity of prokaryotic organisms [40, 42]. There are three reasons for 16 rRNAs as an appropriate marker for taxonomy profiling, and these are (1) the 16 rRNA genes that are present in all prokaryotic organisms; (2) it is almost impossible to experience lateral gene transfer; and (3) the conservative ribosomal protein structure makes the sequence very sustainable [40].

The identity and frequency of microorganisms can be seen by reading 16S rRNA sequences using sequence homology. Readings of genus and species identities can usually be distinguished at a minimum level of 95% for the genus and 97% for species; whereas for strain levels, it is distinguished at a minimum level of 99% [43]. Generally, the V2–V3 region is an excellent area to be used as a gene marker in metagenomic studies. However, several researchers have used various target areas in the V region of the

16S rRNA gene in the analysis of the diversity of microorganisms. According to Zhang et al. [44], the use of different target areas V will result in different bacterial community compositions. The best results were found using the V1–V2 and V1–V3 regions.

#### **Bioinformatics method**

The bioinformatics plays a vital role in the metagenomics analysis. Niu et al. [45] explain the role of bioinformatics in metagenomic analysis, for example, as in the analysis of 16S rRNA data. Analysis of 16S rRNA data can be used to determine the diversity of samples and predict the metabolic pathways of microbes in the sample. An example of a tool used for diversity analysis is MOTHUR. Mallick et al. [46] have reported the use of 16S rRNA sequence data to predict the metabolic pathway of a community from the sample used using the PICRUST software.

The use of bioinformatics tools is based on objectives in exploration. Bioinformatic tools help to analyze environmental samples. Several types of bioinformatics analysis in metagenomics approaches are:

1. Assembly

Assembly is a process of reconstructing short metagenome reads joined to form a long sequence. The long sequence is called as contigs [3]. Assembly uses one of two methods that are often used, OLC and the de Bruijn graph [14]. In addition, other researchers have also developed assembly methods such as hybrid and Iterative joining [47, 48]. However, the de Bruijn graph is the most popular method. The advantage of the de Bruijn graph is cheaper than OLC because it can be built without pairwise comparisons [14]. Bioinformatic tools that can be used in assembly are BBAP, Genovo, MegaGT, and MEGAHIT [49].

2. Binning

Binning is the clustering process of sequences that have been constructed in the assembly process. Binning groups sequences called contigs into classes, so they represent a biological taxon [14]. This method is carried out after assembling raw sequences reads into contigs [50]. Some software options used for binning analysis are MetaWatt [50] and CONCOCT [51]. MetaWatt has advantages that are higher accuracy than existing methods and easy to use [50]. While CONCOCT has reported by the author, this software has high precision and can group complicated microbial communities [51].

3. Sequence analysis

Sequence analysis is a method of finding parts of the same biological sequence [52]. Sequence analysis is divided into two ways, namely, simple alignment and multiple alignments. Simple alignment is the alignment between two sequences, while multiple alignments are the alignment of more than two sequences [53]. One of the tools used for alignment is BLAST (Basic Local Alignment Tool). BLAST is a tool used to compare sequences of various types of organisms. The score of each alignment is given an expectation value ( $E$  value), which is a measure of statistical significance [54].

#### 4. Pfam analysis

Pfam is a database of protein families. Pfam's analysis refers to the double alignment produced using the hidden Markov model. The purpose of Pfam's analysis is to look at the relationship between protein sequences at the family level [54].

#### 5. Analysis of protein structure prediction

The sequence of amino acids is called the primary structure. The primary structure comes from the sequence of the genes that encode it. The structure of proteins is classified as secondary, tertiary, and quaternary structures. Knowledge of the protein structure is fundamental in understanding the function of proteins. Prediction analysis of protein structure by bioinformatics can help in understanding the physical characteristics of a protein and its functions [52].

#### 6. Phylogenetic analysis

Phylogenetic analysis of functional metagenomics refers to procedures used to reconstruct the evolutionary relationships between groups of protein molecules and to predict certain features of a molecule. The methods for forming phylogenetic trees are likelihood methods, parsimony methods, and distance methods. There is no perfect method, and each has specific strengths and weaknesses. The example tools used in phylogenetic analysis are MEGA (Molecular Evolutionary Genetics Analysis), MOLPHY, and PHYLIP [55].

#### Exploration of novel enzymes with the metagenomic approach

The effort to explore natural resources is a strategy in optimizing the use of genetic resources. Enzymes are one of the compounds explored from nature to be taken advantage of in the industrial field. These biocatalysts are not only crucial for cell biochemical processes but also in today's modern industry application. Robinson [2] added that enzymes could be useful in the

pharmaceutical industry for modifying antibiotics, the soap industry, and also for the benefit of forensic and clinical testing.

Exploration of enzymes using a metagenomic approach is not something new. In 1985, Pace and colleagues introduced direct cloning from environmental samples [16]. The first study of screening based on functional genes was successfully conducted by Healy et al. [56], who reported on the isolation of functional genes that encode cellulase enzymes from the environment. Five years later, Rondon et al. [57] have used Bacterial Artificial Chromosome (BAC) as a vector to create a metagenome library from soil samples. Some enzymes are found by Rondon et al. [57], namely, lipases, amylases, and nucleases.

#### Recent research on exploration of novel enzymes with metagenomic approach

Research on enzyme exploration in the past still used conventional methods by culturing the microorganisms on the growth media. However, the development of technology currently directs researchers to the exploration of novel enzymes without culturing on growth media. Cellulase, lipase, and protease enzymes are types of enzymes that are important for industrial processes [58].

#### Cellulases

Cellulases are a group of enzymes catalyzing cellulose polymers into simpler sugars [59]. This enzyme is useful for the paper industry, cotton processing, and detergents [60]. Exploration of cellulase enzymes in a conventional way has placed *Aspergillus* sp. as an organism that has high cellulase activity [61]. However, metagenomic methods reveal that cellulase enzymes can be found widely in various types of organisms. Cui et al. [17] reported that organisms such as *Cloacibacterium*, *Paludibacter*, *Exiguobacterium*, *Acetivibrio*, *Tolomonas*, and *Clostridium* are known to be cellulolytic microbes and have the potential to produce cellulase enzymes. These six genera were found in high cellulose environments in bamboo paper making plants.

Previous research revealed that the genes encoding the cellulase enzyme were also found in the human intestinal microbial community [6] and the microbial community of bagasse waste [7]. Currently, cellulase enzyme exploration also leads to high-temperature environments such as hot springs [62]. The selection of extreme environments is carried out with the hope of getting enzymes with high temperature (thermostable) resistance characteristics.

#### Proteases

Proteases are enzymes that hydrolyze peptide bonds in amino acid chains. This enzyme is used in the detergent, pharmaceutical, and food and beverage industries [63]. Protease sources are spread in several organisms, such

as plants, animals, and microorganisms. Currently, the best-known protease producer in the industry is *Bacillus* sp. [64]. The development of metagenomic technology enables the search for other organisms that are potentially more efficient in terms of effectiveness. Biver et al. [18] reported the discovery of a new protease-coding gene derived from a microbe similar to *Desulfobacter postgatei* 2 ac9 with a similarity rate of 69%. Also, Devi et al. [65] reported the findings of the Prt1A gene that encodes the protease enzyme from organic sludge. The protease enzyme from the Prt1A gene is known to be optimal at 55 °C. The following year, Pessoa et al. [66] discovered a gene that codes for proteases with optimum activity at 60 °C.

### Lipases

Lipases are enzymes that catalyze the hydrolytic cleavage of the ester bonds between carboxylic acids and alcohol groups [67]. This enzyme is used in the detergent, food, biodiesel, and bioremediation industries. *Bacillus* spp. bacteria such as *B. alcalophilus*, *B. licheniformis*, *B. pumilus*, and *B. subtilis* are the most well-known producers of bacterial lipases at present [30, 31, 68].

Researchers are currently competing in exploring other organisms that have the potential to produce better lipase. Hardeman & Sjoling [69], with a functional metagenomic approach, have found the h1Lip1 gene that has a similarity to the lipase of *Pseudomonas putida* with a similarity level of 54%. Lipase enzyme is from the optimum h1Lip1 gene at 35 °C (low temperature). According to López-lópez et al. [67], the maximum lipase character at low temperatures is generally suitable for the cold washing process in detergents. Besides, many other enzymes sourced from the metagenome library have unique biochemical properties that make them valuable for industrial applications. An example is an enzyme that is resistant to solvents, and

detergents found in soils are contaminated with petroleum hydrocarbons [19].

### Other enzymes

There are many enzymes found by researchers from metagenome-source, which can potentially be commercialized. Recently, Sharma et al. [70] have reported novel bleomycin resistance dioxygenase (BRPD) from contaminated agricultural soil. It has a function in the bioremediation process by catalyzing the degradation of hydrocarbon substrate like pesticides. In addition, Berini et al. [71] also have found *53D1* gene encoding chitinases which can potentially be used for controlling plant pests. They investigated chitinases controlling in *Bombyx mori*, a Lepidoptera. The result showed that chitinase (*53D1* gene) was a promising enzyme used as an insecticide. Other recent studies have also revealed enzymes derived from metagenome sources, such as oxoflavin-degrading enzyme used in the agricultural industry [72], transaminases used in the pharmaceutical industry [73], and AHL-lactonase [74].

### The patented enzymes from metagenome-source for industry

In the past 5 years, several patented enzymes beneficial to the industry are published. Previously, the patented enzymes for commercialization are reported by Berini et al. [75]. This review presents novel patented enzymes published in the past 5 years (Table 3). The lists patented enzymes from metagenome source include cellulases, protease, lipase,  $\alpha$ -amylase, chitinase,  $\beta$ -glucosidase, and endoglucanase.

### Challenges in exploration of novel enzymes with the metagenomic approach

The selection of sampling locations is a challenge for researchers to explore novel enzymes. A location will determine the role of enzyme characterization. Each

**Table 3** Examples of patented enzymes from metagenome source in the past 5 years

Patent Number	Country	Assignee	Enzyme	Source	Year	Application
CN108463551A	China	Scientific and Industrial Research Council	Cellulases	Soil	2016	food and feed industry, detergent, weaving, and biofuel industry
BR102016000771A2	Brazil	Universidade Estadual de Santa Cruz	Proteases	mangrove sediment	2016	antitumor, antifungal, antiviral, and antiparasitic treatment
KR102026836B1	South Korea	Korea Research Institute of Chemical Technology	Lipase	Soil	2018	industrial mass-production, biopharmaceuticals, and biodiesel
KR101816615B1	South Korea	Republic of Korea	$\alpha$ -Amylase	black goat rumen	2015	feed additive, detergents, and biofuel
CN107475273B	China	Chengdu Institute of Biology, Chinese Academy of Sciences	Chitinase	Wetland environment	2017	the industries of food, medicine, agriculture, and cosmetics
CN107828806A	China	Guangdong Pharmaceutical University	$\beta$ -glucosidase	Soil	2017	The industry of pharmaceuticals, food, bioethanol, and medicine
JP6552098B2	Japan	Honda Motor Co., Ltd., Kazusa DNA Research Institute	Endoglucanase	Hot spring soil	2016	bioethanol

location has a unique ecological niche for the exploration of novel enzymes. Unique niches are created by functional interactions between the microbial community and their environment [76, 77]. Mhuantong et al. [76] reported the discovery of high cellulolytic bacteria in sugarcane bagasse samples. The number of cellulolytic bacteria even looks significant when compared to in-cellulolytic bacteria. Nie et al. [78] also have reported that the microbial community in the oil environment had genes that encoded hydrocarbon degradation enzymes. The existence of hydrocarbon degradation enzymes can not be separated from the oil environment rich in hydrocarbon compounds. The research evidence reinforces the theory that the environment determines the characterization of enzymes, so the consideration of selecting the sampling location needs to be adjusted to the type of enzyme to be explored. Also, the characteristics of sampling locations pose challenges for researchers. Locations with extreme characters require special techniques and caution in sampling, for example, sampling at hot spring locations [79].

The choice of DNA extraction methods from environmental samples is also a challenge because it has greater difficulty than DNA extraction from a single genome. Difficulties arise due to DNA from the environment that contains more contaminants, such as humic acid, protein, and carbohydrates. In addition, if clay sample is used, it will be more difficult to extract the DNA as it is bound to soil particles [80]. Therefore, researchers need to do the special treatment of samples that have unique characteristics.

## Conclusions

Metagenomics is the study of genomes from microorganism communities in the environment. Metagenomics is divided into two primary studies, namely, structural and functional metagenomics. Basic structural metagenomics method consists of assembly, binning, and microbial community analysis such as taxonomic profiling, gene prediction, and metabolic pathways. Meanwhile, functional metagenomics approach consists of gene construction, screening, heterologous overexpression, bioinformatic analysis, and protein product characterization.

Exploration of novel enzymes with a metagenomic approach has revealed several novel enzymes from nature, such as cellulases, proteases, lipases, and other enzymes such as BRPD, chitinases, oxoflavin-degrading enzyme, transaminases, and AHL-lactonase. The existence of metagenomics has helped researchers uncover novel enzymes from nature that are beneficial to the industries. Understanding of metagenomic and its application is expected to have an impact on the development of technology that is useful for humanity.

## Abbreviations

eDNA: Environmental DNA; NGS: Next Generation Sequencing; bp: Base pair; kb: Kilobase pair; Gb: Gigabase pair; rRNA: Ribosome-Ribonucleic Acid; LSU: Large Subunits; SSU: Small Subunits; BLAST: Basic Local Alignment Tool; MEGA: Molecular Evolutionary Genetics Analysis; BAC: Bacterial Artificial Chromosome

## Acknowledgements

Not applicable

## Authors' contributions

FAP carried out literature search, survey and writing manuscript. HPK, WW, AS and NN worked to edit and shape the manuscript. AB designed, edit and writing the manuscript. All authors contributed to the final manuscript.

## Funding

AB and FAP would like to thank Faculty of Science and Mathematics-Diponegoro University - Indonesia for the research grant under the scheme of PNBP.

## Availability of data and materials

Not applicable

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Biology, Faculty of Science and Mathematics, Diponegoro University, Semarang City 50275, Indonesia. <sup>2</sup>Biotechnology Study Program, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Sudharto SH, Semarang 50275, Indonesia. <sup>3</sup>Molecular and Applied Microbiology Laboratory, Center Central Laboratory of Research and Service - Diponegoro University, Jl. Prof. Sudharto SH, Semarang 50275, Indonesia.

Received: 15 January 2020 Accepted: 10 June 2020

Published online: 04 August 2020

## References

1. Jansson JK, Hofmockel KS (2018) ScienceDirect The soil microbiome — from metagenomics to metaphenomics. *Curr Opin Microbiol* 43:162–168 <https://doi.org/10.1016/j.mib.2018.01.013>
2. Robinson PK (2015) Enzymes: principles and biotechnological applications. *Essays Biochem* 59:1–41 <https://doi.org/10.1042/BSE0590001>
3. Papudeshi B, Haggerty JM, Doane M, Morris MM, Walsh K, Beattie DT, Pande D, Zaeri P, Silva GGZ, Thompson F, Edwards RA, Dinsdale EA (2017) Optimizing and evaluating the reconstruction of Metagenome-assembled microbial genomes. *BMC Genomics* 18:1–13 <https://doi.org/10.1186/s12864-017-4294-1>
4. Neelakanta G, Sultana H (2013) The use of metagenomic approaches to analyze changes in microbial communities. *Microbiol Insights* 6:37–48 <https://doi.org/10.4137/MBIS10819>
5. Tripathi M, Singh DN, Vikram S, Singh VS, Kumar S (2018) Metagenomic approach towards bioprospection of novel biomolecule(s) and environmental bioremediation. *Ann Res Rev Biol* 22:1–12 <https://doi.org/10.9734/ARRB/2018/38385>
6. Aggarwala V, Liang G, Bushman FD (2017) Viral communities of the human gut : metagenomic analysis of composition and dynamics. *Mob DNA* 8:1–10 <https://doi.org/10.1186/s13100-017-0095-y>
7. Kanokratana P, Mhuantong W, Laothanachareon T, Tangphatsornruang S, Eurwilaichitr L, Pootanakit K, Champreda V (2013) Phylogenetic analysis and metabolic potential of microbial communities in an industrial bagasse collection site. *Environ Microbiol* 66:322–334 <https://doi.org/10.1007/s00248-013-0209-0>
8. Jacob JH, Hussein EI, Shakhathreh MAK, Cornelison CT (2017) Microbial community analysis of the hypersaline water of the Dead Sea using high-



- throughput amplicon sequencing. *MicrobiologyOpen* 6:1–6 <https://doi.org/10.1002/mbo3.500>
9. Gilbert J (2013) Metagenomics , metadata , and meta-analysis meta-analysis. *Encyclopedia Metagenomics:1–4* <https://doi.org/10.1007/978-1-4614-6418-1>
  10. Nazir A (2016) Review on metagenomics and its applications. *IJIR* 2:277–286
  11. Ghosh A, Mehta A, Khan AM (2018) Metagenomic analysis and its applications. In: *Encyclopedia of Bioinformatics and Computational Biology*, 1st edn. Elsevier Inc, Amsterdam, pp 1–10 <https://doi.org/10.1016/B978-0-12-809633-8.20178-7>
  12. Alves LDF, Westmann CA, Lovate GL, Marcelino G, De Siqueira V, Borelli TC, Guazzaroni M (2018) Metagenomic Approaches for understanding new concepts in microbial science. *Int J Genomics* 2018:1–15 <https://doi.org/10.1155/2018/2312987>
  13. Jimenez DJ, Andreote FD, Chaves D, Montan JS, Osorio-Forero C, Junca H, Zambrano MM, Baena S (2012) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. *PLoS One* 7:1–15 <https://doi.org/10.1371/journal.pone.0052069>
  14. Roumpeka DD, Wallace RJ, Escalettes F, Fotheringham I, Watson M (2017) A review of bioinformatics tools for bio-prospecting from metagenomic sequence data. *Front Genet* 8:1–10 <https://doi.org/10.3389/fgene.2017.00023>
  15. Edet UO, Antai SP, Brooks AA, Asitok AD, Enya O (2017) An overview of cultural, molecular and metagenomic techniques in description of microbial diversity. *J Adv Microbiol* 7:1–19 <https://doi.org/10.9734/JAMB/2017/37951>
  16. Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. *77: 1153–1161* <https://doi.org/10.1128/AEM.02345-10>
  17. Cui J, Mai G, Wang Z, Liu Q, Zhou Y, Ma Y (2019) Metagenomic insights into a cellulose-rich niche reveal microbial cooperation in cellulose degradation. *Front Microbiol* 10:1–12 <https://doi.org/10.3389/fmicb.2019.00618>
  18. Biver S, Portetelle D, Vandenberg M (2013) Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. *SpringerPlus* 2: 1–10 <https://doi.org/10.1186/2193-1801-2-410>
  19. Pereira MR, Mercaldi GF, Maester TC, Balan A, & Lemos EG de M (2015) Est16, a New esterase isolated from a metagenomic library of a microbial consortium specializing in diesel oil degradation. *PLoS One* 10: 1–16. <https://doi.org/10.1371/journal.pone.0133723>
  20. Liu X, Kokare C (2017) Microbial enzymes of use in industry. In: *Brahmachari G (ed) Biotechnology of Microbial Enzymes*. Elsevier Inc., London, pp 267–298 <https://doi.org/10.1016/B978-0-12-803725-6.00011-X>
  21. National Research Council (2007) *The new science of metagenomics: revealing the secrets of our microbial planet*. The National Academies Press
  22. Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci* 74:5088–5090
  23. Escobar-zepeda A, De León AV, Sanchez-flores A (2015) The road to metagenomics : from microbiology to DNA sequencing technologies and bioinformatics. *Front Genet* 6:1–15 <https://doi.org/10.3389/fgene.2015.00348>
  24. Kchouk M, Gibrat J, Elloumi M (2017) Generations of sequencing technologies: from first to next generation. *Biol Med* 9:1–8 <https://doi.org/10.4172/0974-8369.1000395>
  25. Bragg L, Tyson GW (2014) *Metagenomics Using next-generation sequencing*. In: *Environmental Microbiology: Methods and Protocols* (Vol. 1096). Springer Science, New York <https://doi.org/10.1007/978-1-62703-712-9>
  26. Pucker B, Schilbert HM, Schumacher SF (2019) Integrating molecular biology and bioinformatics education. *J Integ Bioinform* 16:1–7 <https://doi.org/10.1515/jib-2019-0005>
  27. Tanveer A, Yadav S, Yadav D (2016) Comparative assessment of methods for metagenomic DNA isolation from soils of different crop growing fields. *3. Biotech* 6:1–5 <https://doi.org/10.1007/s13205-016-0543-2>
  28. Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckley TR, Cruickshank R, Dopheide A, Handley KM, Hermans S, Kamke J, Lee CK, MacDiarmid R, Morales SE, Orlovich DA, Smissen R, Wood J, Holdaway R (2018) Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *N Z J Ecol* 42:1–50 <https://doi.org/10.20417/nzjecol.42.9>
  29. Gutiérrez-lópez R, Puente JM, Gangoso L, Soriguer RC, Figuerola J (2014) Comparison of manual and semi-automatic DNA extraction protocols for the barcoding characterization of hematophagous louse flies (Diptera : Hippoboscidae). *J Vector Ecol* 40(1):11–15 <https://doi.org/10.1111/jvec.12127>
  30. Hassan M, Essam T, Megahed S (2018b) Illumina sequencing and assessment of new cost-efficient protocol for metagenomic-DNA extraction from environmental water samples. *Braz J Microbiol* 49:1–8 <https://doi.org/10.1016/j.bjm.2018.03.002>
  31. Hassan SWM, El Latif HHA, Ali SM (2018a) Production of cold-active lipase by free and immobilized marine bacillus cereus HSS : application in wastewater treatment. *Front Microbiol* 9:1–13 <https://doi.org/10.3389/fmicb.2018.02377>
  32. Felczykowska A, Krajewska A, Zielińska S, Łoś JM (2015) Sampling , metadata and DNA extraction — important steps in metagenomic studies. *62(1)*
  33. Young JM, Rawlence NJ, Weyrich LS, Cooper A (2014) Limitations and recommendations for successful DNA extraction from forensic soil samples : a review. *Sci Justice* 54:1–7 <https://doi.org/10.1016/j.scjus.2014.02.006>
  34. Li X, Wu Y, Zhang L, Cao Y, Li Y, Li J, Zhu L, Wu G (2014) Comparison of three common DNA concentration measurement methods. *Anal Biochem* 451:18–24 <https://doi.org/10.1016/j.ab.2014.01.016>
  35. Sun W (2010) *General Procedures. In molecular diagnostics: techniques and applications for the clinical laboratory* (p. 92675). London: Academic Press. <https://doi.org/10.1016/C2009-0-00182-2>
  36. Lucena-aguilár G, Sanchez-Lopez AM, Barberan-Aceituno C, Carrillo-Avila JA, Lopez-Guerrero JA, Aguilar-Quesada R (2016) DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreserv Biobanking* 14:264–270 <https://doi.org/10.1089/bio.2015.0064>
  37. Barril P, Nates S (2012) Introduction to agarose and polyacrylamide gel electrophoresis matrices with respect to their detection sensitivities. In: *Gel Electrophoresis – Principles and Basics*. IntechOpen, Rijeka, pp 3–14
  38. Yilmaz M, Ozic C, Gok İ (2012) Principles of nucleic acid separation by agarose gel electrophoresis. In: *Gel Electrophoresis – Principles and Basics*. IntechOpen, Rijeka, pp 33–40
  39. Gutiérrez-lucas LR, Montor-antonio JJ, Cortés-lópez NG, Moral S (2014) Strategies for the Extraction , Purification and amplification of metagenomic DNA from soil growing sugarcane. *Adv Biol Chem* 4:281–289 <https://doi.org/10.4236/abc.2014.44034> Strategies
  40. Ramazzotti M, & Bacci G (2018) 16S rRNA-based taxonomy profiling in the metagenomics era. In *Metagenomics* (pp. 103–119). Elsevier Inc. <https://doi.org/10.1016/B978-0-08-102268-9.00005-7>
  41. Singer E, Bushnell B, Coleman-derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng J, Copeland A, Klenk H, Hallam SJ, Hugenholtz P, Tringe SG, Woyle T (2016) High-resolution phylogenetic microbial community profiling. *ISME J* 10:2020–2032 <https://doi.org/10.1038/ismej.2015.249>
  42. Nguyen N, Warnow T, Pop M, White B (2016) A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms Microbiomes* 2:1–8 <https://doi.org/10.1038/npjbio>
  43. Del Chierico F, Ancora M, Marcacci M, Cammà C, Putignani L, Conti S (2015) Choice of next-generation sequencing pipelines. In: *Bacterial Pangenomics*, vol 1231. Humana Press, New York, pp 31–47 <https://doi.org/10.1007/978-1-4939-1720-4>
  44. Zhang J, Ding X, Guan R, Zhu C, Xu C, Zhu B, Zhang H, Xiong Z (2018) Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. *Sci Total Environ* 618:1254–1267 <https://doi.org/10.1016/j.scitotenv.2017.09.228>
  45. Niu S, Yang J, Mcdermaid A, Zhao J, Kang Y (2018) Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. *Brief Bioinform* 19:1415–1429 <https://doi.org/10.1093/bib/bbx051>
  46. Mallick H, Franzosa EA, McIver LJ, Banerjee S, Sirota-Madi A, Kostic AD, Clish CB, Vlamakis XRJ, Huttenhower C (2019) Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nat Commun* 10:1–11 <https://doi.org/10.1038/s41467-019-10927-1>
  47. Gregor I, Scho A, Mchardy AC (2016) Snowball : strain aware gene assembly of metagenomes. *Bioinformatics* 32:i649–i657 <https://doi.org/10.1093/bioinformatics/btw426>
  48. Ruby JG, Bellare P, & DeRisi JL (2013) PRICE : Software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)* 3: 865–880. <https://doi.org/10.1534/g3.113.005967>
  49. Ayling M, Clark MD, Leggett RM (2019) New approaches for metagenome assembly with short reads. *Brief Bioinform* 00:1–11 <https://doi.org/10.1093/bib/bbz020>
  50. Strous M, Kraft B, Bisdorf R, Tegetmeyer HE (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front Microbiol* 3:1–11 <https://doi.org/10.3389/fmicb.2012.00410>
  51. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by

- coverage and composition. *Nat Methods* 11:1144–1146 <https://doi.org/10.1038/nmeth.3103>
52. Raza K (2012) Application of data mining in bioinformatics. *Indian J Comp Sci Eng* 1:114–118
53. Diniz WJS, Canduri F (2017) Bioinformatics : an overview and its applications. *Genet Mol Res* 16:1–21
54. Khandelwal I, Sharma A, Agrawal PK, Shrivastava R (2017) Bioinformatics Database Resources. In: Library and information services for bioinformatics education and research. IGI Global, Hershey, pp 45–90 <https://doi.org/10.4018/978-1-5225-1871-6.ch004>
55. Mehmood MA, Sehar U, Ahmad N (2014) Use of bioinformatics tools in different spheres of life sciences. *Data Mining Genom Proteomics* 5:1–13 <https://doi.org/10.4172/2153-0602.1000158>
56. Healy FG, Ray RM, Aldrich HC, Wilkie AC, Ingram LO, Shanmugam KT (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* 43:667–674 <https://doi.org/10.1007/BF00164771>
57. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loinacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osborne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome : a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66:2541–2547 <https://doi.org/10.1128/aem.66.6.2541-2547.2000>
58. Bashir Y, Singh SP, Konwar BK (2014) Metagenomics : an application based perspective. *Chin J Biol* 2014:1–7 <https://doi.org/10.1155/2014/146030>
59. Escuder-Rodríguez J, DeCastro ME, Cerdán ME, Rodríguez-Belmonte E, Becerra M, González-Siso MI (2018) Cellulases from thermophiles found by metagenomics. *Microorganisms* 6:1–26. <https://doi.org/10.3390/microorganisms6030066>
60. Tiwari R, Nain L, Labrou NE, Shukla P (2017) Bioprospecting of functional cellulases from metagenome for second generation biofuel production : a review. *Crit Rev Microbiol*:1–14 <https://doi.org/10.1080/1040841X.2017.1337713>
61. Paula CCPDE, Montoya QV, Meirelles LA (2019) High cellulolytic activities in filamentous fungi isolated from an extreme oligotrophic subterranean environment (Catão cave) in Brazil. *An Acad Bras Cienc* 91:1–11 <https://doi.org/10.1590/0001-3765201920180583.ABstract>
62. Schröder C, Elleuche S, Blank S, Antranikian G (2014) Enzyme and microbial technology characterization of a heat-active archaeal Beta-glucosidase from a hydrothermal spring metagenome. *Enzym Microb Technol* 57:48–54 <https://doi.org/10.1016/j.enzmictec.2014.01.010>
63. Tavano OL, Berenguer-murcia A, Secundo F, Fernandez-lafuente R (2018) Biotechnological applications of proteases in food technology. *Compr Rev Food Sci Food Saf* 17:412–436 <https://doi.org/10.1111/1541-4337.12326>
64. Razaq A, Shamsi S, Ali A, Ali Q, Sajjad M, Malik A, Ashraf M (2019) Microbial Proteases Applications. *Front Bioengineering Biotechnol* 7:1–20 <https://doi.org/10.3389/fbioe.2019.00110>
65. Devi SG, Fathima AA, Sanitha M, Iyappan S, Curtis WR, Ramya M (2016) Expression and characterization of alkaline protease from the metagenomic library of tannery activated sludge. *J Biosci Bioeng* 122:694–700 <https://doi.org/10.1016/j.jbiosc.2016.05.012>
66. Pessoa TBA, Rezende R, Marques EDLS, Pirovani CP, Santos TFD, Goncalves ACDS, Romano CC, Dotivo NC, Freitas ACO, Salay LC, Dias JCT (2017) Metagenomic alkaline protease from mangrove sediment. *J Basic Microbiol* 2017:1–12 <https://doi.org/10.1002/jobm.201700159>
67. López-López O, Cerdán ME, Siso MIG (2014) New extremophilic lipases and esterases from metagenomics. *Curr Protein Pept Sci* 15:445–455 <https://doi.org/10.2174/1389203715666140228153801>
68. Kaur G, Singh A, Sharma R, Sharma V, Verma S, Sharma PK (2016) Cloning, expression, purification and characterization of lipase from *Bacillus licheniformis*, isolated from hot spring of Himachal Pradesh, India. *3 Biotech* 6:1–10 <https://doi.org/10.1007/s13205-016-0369-y>
69. Hardeman F, Sjolting S (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol Ecol* 59:524–534 <https://doi.org/10.1111/j.1574-6941.2006.00206.x>
70. Sharma V, Kumar R, Sharma VK, Yadav AK, Tirola M, Sharma PK (2019) Expression, purification, characterization and in silico analysis of newly isolated hydrocarbon degrading bleomycin resistance dioxygenase. *Mol Biol Rep* 47:533–544 <https://doi.org/10.1007/s11033-019-05159-x>
71. Berini F, Casartelli M, Montali A, Reguzzoni M, Tettamanti G, Marinelli F (2019) Metagenome-sourced microbial chitinases as potential insecticide proteins. *Front Microbiol* 10:1–12 <https://doi.org/10.3389/fmicb.2019.01358>
72. Choi J, Nguyen CM, Lee B, Park JH, Oh JY, Choi JS, Kim J, Song JK (2018) Isolation and characterization of a novel metagenomic enzyme capable of degrading bacterial phytotoxin toxoflavin. *PLoS One* 13:1–14 <https://doi.org/10.1371/journal.pone.0183893>
73. Leipold L, Dobrijevic D, Jeffrie JWE, Bawn M, Moody TS, Ward JM, Hailes HC (2019) The identification and use of robust transaminases from a domestic drain metagenome. *Green Chem* 21:75–86 <https://doi.org/10.1039/c8gc02986e>
74. Fan X, Liang M, Wang L, Chen R, Li H, Liu X (2017) Aii810, a novel cold-adapted N -acylhomoserine lactonase discovered in a metagenome , can strongly attenuate *Pseudomonas aeruginosa* virulence factors and biofilm formation. *Front Microbiol* 8(1950):1–11. <https://doi.org/10.3389/fmicb.2017.01950>
75. Berini F, Casciello C, Marcone GL, Marinelli F (2017) Metagenomics : novel enzymes from non-culturable microbes. *FEMS Microbiol Lett* 364:1–19 <https://doi.org/10.1093/femsle/fnx211>
76. Mhuantong W, Charoensawan V, Kanokratana P, Tangphatsomruang S (2015) Comparative analysis of sugarcane bagasse metagenome reveals unique and conserved biomass-degrading enzymes among lignocellulolytic microbial communities. *Biotechnol Biofuels* 8:1–17 <https://doi.org/10.1186/s13068-015-0200-8>
77. Staley C, Sadowsky MJ (2018) Practical considerations for sampling and data analysis in contemporary metagenomics-based environmental studies. *J Microbiol Methods* 154:14–18 <https://doi.org/10.1016/j.jmimet.2018.09.020>
78. Nie Y, Zhao J, Tang Y, Guo P, Yang Y, Wu X, Hickey WJ (2016) Species divergence vs. functional convergence characterizes crude oil microbial community assembly. *Front. Microbiol* 7:1–11 <https://doi.org/10.3389/fmicb.2016.01254>
79. Saxena R, Dhakan DB, Mittal P, Waiker P (2017) Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front Microbiol* 7: 1–17 <https://doi.org/10.3389/fmicb.2016.02123>
80. Zaveri P, Patel R, Patel M, Sarodia D, Munshi NS (2017) Modification of extraction method for community DNA isolation from salt affected compact wasteland soil samples. *MethodsX* 4:63–67 <https://doi.org/10.1016/j.mex.2017.01.002>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)